

DISCRIMINAÇÃO ALGORÍTMICA NO BRASIL: UMA ANÁLISE DA PESQUISA JURÍDICA E SUAS PERSPECTIVAS PARA A COMPREENSÃO DO FENÔMENO

ALGORITHMIC DISCRIMINATION IN BRAZIL: AN ANALYSIS OF LEGAL RESEARCH AND PERSPECTIVES FOR UNDERSTANDING THE PHENOMENON

NILTON SAINZ¹

Pontifícia Universidade Católica do Paraná (PUCPR). Curitiba (PR). Brasil.

EMERSON GABARDO^{2,1}

I Pontifícia Universidade Católica do Paraná (PUCPR). Curitiba (PR). Brasil.

NATÁLIA ONGARATTO^{3,II}

II Fundação Escola do Ministério Público do Estado do Paraná (FEMPAR). Curitiba (PR). Brasil.

RESUMO: Este artigo tem como objetivo analisar a abordagem da pesquisa acadêmica sobre a discriminação algorítmica no campo do Direito no Brasil, fornecendo uma descrição do estágio atual dessa discussão. Com base em uma revisão de escopo da literatura jurídica, foram identificadas as áreas de Direitos Humanos e Direito Administrativo como aquelas que possuem as principais concentrações de estudos sobre o tema. Por outro lado, subáreas do Direito, como Direito Civil, Direito do Trabalho e Direito Processual Penal, estão em desenvolvimento. A pesquisa foi realizada por meio de uma busca sistemática em periódicos indexados na *Scopus*, *Web of Science*, *SciELO* e *Google Scholar*, além de publicações no Congresso Internacional de Direito e Inteligência Artificial. A hipótese inicial de que a pesquisa sobre a discriminação algorítmica no Direito no Brasil estaria em estágio inicial foi confirmada, porém com identificação de forte potencial de crescimento. As conclusões também apontaram, na área do Direito, uma concentração de estudos em casos discriminatórios semelhantes, bem como a identificação de objetos de análise e problemáticas similares. A pesquisa e suas conclusões se justificam considerando-se a importância de abordagens interdisciplinares a fim de ser obtida uma compreensão abrangente e atualizada do fenômeno da discriminação algorítmica, possibilitando o desenvolvimento de soluções regulatórias mais eficientes para as novas tecnologias.

PALAVRAS-CHAVE: Discriminação algorítmica; Inteligência artificial; Direitos Humanos; Direito Digital; Novas tecnologias.

ABSTRACT: This article aims to analyze the approach of academic research on algorithmic discrimination in the field of Law in Brazil, providing a description of the current stage of this discussion. Based on a scope review of legal literature, the areas of Human Rights and Administrative Law were identified as having the main concentrations of studies on the subject. On the other hand, sub-areas of Law such as Civil Law, Labor Law, and Criminal Procedural Law are under development. The research was conducted through a systematic search in journals indexed in Scopus, Web of Science, SciELO, and Google Scholar, as well as publications from the International Congress on Law and Artificial Intelligence. The initial hypothesis that research on algorithmic discrimination in Law in Brazil is in an early stage was confirmed, but with the identification of strong growth

¹ <https://orcid.org/0000-0002-3957-2714>

² <https://orcid.org/0000-0002-1798-526X>

³ <https://orcid.org/0009-0004-0902-0064>



potential. The conclusions also indicated that there is a concentration of studies in the field of Law on similar discriminatory cases, as well as the identification of similar objects of analysis and issues. The research and its conclusions are justified considering the importance of interdisciplinary approaches in order to obtain a comprehensive and up-to-date understanding of the phenomenon of algorithmic discrimination, enabling the development of more efficient regulatory solutions for new technologies.

KEYWORDS: Algorithmic Discrimination; Artificial Intelligence; Human Rights; Digital Law; New technologies.

INTRODUÇÃO

Em 2019, um caso de discriminação de gênero por meio da ferramenta de pontuação de crédito do *Apple Card* viralizou nas redes sociais e desencadeou uma investigação regulatória das práticas adotadas pelo *software* utilizado na empresa. As vítimas denunciaram a recusa de concessão de crédito à esposa, mesmo ela possuindo melhores atributos de crédito do que seu marido, o que chamou a atenção e gerou ampla repercussão (VIGDOR, 2019). Após o caso COMPAS ocorrido nos Estados Unidos em 2016, a questão da discriminação algorítmica já havia ganhado maior notoriedade em diversos setores do sistema judiciário. Em 2022, um novo sistema de pontuação de riscos foi divulgado, denominado *Prisoner Assessment Tool Targeting Estimated Risk and Needs* (PATTERN), o qual visava estimar a probabilidade de reincidência de um detento após a sua liberação. Análises do sistema revelaram que, mais uma vez, um algoritmo superestimava a reincidência de indivíduos não brancos na prática de novos crimes (PURVES; DAVIS, 2022). Trata-se, portanto, de um problema persistente e atual. O avanço tecnológico na sociedade não apenas traz consigo transformações positivas nos costumes, também sendo capaz de reinventar problemas sociais que lhe são anteriores, agora em um novo, e ainda mais impactante, contexto (REYNA; GABARDO; SANTOS, 2020).

No campo das Ciências Jurídicas, termos como inteligência artificial (IA) e algoritmos estão cada vez mais presentes no cotidiano profissional. No entanto, não há um amplo consenso sobre seus significados, o que torna a síntese desses novos signos por si só desafiadora. Em um breve esforço, a IA pode ser definida como um sistema de ação racional projetado para aprender e se adaptar com base em dados e experiências, com o objetivo de reconhecer padrões, fazer previsões ou tomar decisões. Os algoritmos, por sua vez, podem ser compreendidos como os mecanismos responsáveis por executar as tarefas das IAs, ou seja, eles seguem um

encadeamento lógico preciso com base nos dados e informações (*inputs*) fornecidos (RUSSELL; NORVIG, 2021).

A discriminação negativa (ilegítima), por si mesma, é um fenômeno que ultrapassa o mundo digital. O racismo estrutural vivenciado na história brasileira é um exemplo eloquente e que impacta na hermenêutica jurídica (SANTOS, 2021b). A discriminação algorítmica, por sua vez, pode decorrer tanto do resultado da programação efetuada para o *software* quanto ser uma reprodução de preconceitos e discriminações já existentes na sociedade, e que adquirem um potencial de perpetuação de vieses por meio de diferentes tratamentos de dados (BAROCAS; SELBST, 2016). Ainda não existe uma classificação consensual do conceito de discriminação algorítmica. Uma proposição inovadora é de Mendes e Mattiuzzo (2019, p. 14-15) que apresentam quatro tipos de discriminação que ocorrem a partir da utilização de algoritmos: I) discriminação por erro estatístico; II) discriminação por generalização; III) discriminação por uso de informações sensíveis; IV) e, por fim, discriminação limitadora do exercício de direitos. Embora possuam semelhanças, esses tipos de discriminação algorítmica se distinguem principalmente pelos elementos causais. Em todos os casos, a discriminação efetuada é relevante para o Direito, notadamente nos seus aspectos de regulação e controle das novas tecnologias.

A discriminação por erro estatístico ocorre quando há falhas na coleta de dados ou na construção do código do algoritmo. Um exemplo disso pode ser um erro amostral na população presente nos dados, o que resulta na subestimação ou desconsideração de certos grupos sociais. Além disso, pode haver um desequilíbrio no cálculo estatístico, atribuindo um valor maior a uma determinada característica. Por outro lado, a discriminação por generalização decorre da limitação das modelagens estatísticas. Embora não haja falhas no cálculo, esses modelos, que estão associados ou correlacionados com características específicas, não conseguem considerar a diversidade individual de uma população, ignorando a "margem de erro". Já a discriminação por uso de informações sensíveis é baseada na utilização de características intrínsecas a indivíduos ou grupos historicamente discriminados. Esse tipo de discriminação pode ocorrer por violações de dados, uma vez que certas informações são protegidas por lei – ou, ao contrário, pode desconsiderar estas variáveis quando elas são essenciais para a obtenção de um resultado justo. Por fim, a discriminação limitadora do exercício de direitos é caracterizada pela forma como os dados são utilizados, ou seja, não é resultado de um erro estatístico ou da

qualidade da informação, mas sim do impacto negativo que o resultado do algoritmo tem sobre os direitos de um indivíduo.

A Resolução nº 332/2020 do Conselho Nacional de Justiça, em seu art. 7º, destaca o princípio da não discriminação como fundamental para a implementação de ferramentas de IA no sistema de justiça brasileiro. Mobilizações recentes, como as da Justiça do Trabalho e do Ministério Público do Trabalho (MPT), demonstram que a discriminação algorítmica é uma pauta presente nesses órgãos. Em 2020, o MPT criou o grupo de estudo Diversidade e Tecnologia, com o objetivo de qualificar seus membros e disseminar informações sobre discriminação algorítmica e tecnológica. Como resultado desse esforço, o grupo publicou a coletânea “O uso de dados pessoais e inteligência artificial na relação de trabalho: proteção, discriminação, violência e assédio digital” (ARAÚJO *et al.*, 2022), que inclui capítulos abordando as múltiplas faces da discriminação algorítmica nas relações de trabalho. O ponto central dessas reflexões tem sido questionar quem são os beneficiados por tecnologias discriminatórias e quais grupos são mais vulneráveis e excluídos dos processos de desenvolvimento e aplicação dessas novas tecnologias nas relações sociais em geral.

Motivado pela relevância da problemática jurídica e social, o objetivo deste artigo é analisar como a discriminação algorítmica tem sido abordada na pesquisa acadêmica na área do Direito no Brasil, bem como fornecer uma descrição do estágio atual dessa discussão. Portanto, busca-se compreender como esse fenômeno tem ganhado importância nos estudos jurídicos recentes, examinando os objetos de análise, as subáreas de concentração dos estudos, o mapeamento de casos de discriminação citados, as principais problematizações sobre a discriminação algorítmica e as possibilidades futuras de avanços nessa temática. A hipótese deste estudo parte do pressuposto de que o tema está sendo desenvolvido de forma concentrada em duas áreas principais: Direitos Humanos e Direito Administrativo, recebendo pouca atenção de áreas subjacentes ao fenômeno da discriminação algorítmica. Isso resulta em uma concentração de abordagens, casos discriminatórios e problematizações em subáreas específicas.

Apoiando-se em abordagens contemporâneas de investigações sobre a literatura jurídica (DANCEANU, 2019; DINOS *et al.*, 2015; HORÁKA; LACKOB; KLOCEKB, 2021), optou-se pelo método de revisão de escopo para a realização deste estudo, utilizando uma busca sistemática de literatura. Essa escolha metodológica tem como objetivo reduzir os vieses na

seleção do que é considerado para uma revisão narrativa de literatura e mitigar os efeitos na interpretação do fenômeno por meio de uma amostragem significativa de publicações sobre a temática. A coleta dessa literatura foi baseada em periódicos indexados na *Scopus*, *Web of Science* (WoS) e *SciELO*, além da inclusão de artigos indexados no *Google Scholar* (GS) publicados em periódicos de estrato superior no Qualis-Capes e *papers* publicados no Congresso Internacional de Direito e Inteligência Artificial (CIDIA).

Dito isto, após a introdução dos conceitos, problemáticas e metodologia de pesquisa, este artigo está dividido em três seções. A primeira aborda os detalhes metodológicos, começando com a apresentação dos princípios da revisão de escopo, onde são detalhados os procedimentos adotados para a análise da literatura, seguindo um protocolo adaptado. Em seguida, são discutidos os critérios utilizados na busca sistemática. A segunda seção do artigo é dividida em duas partes: a primeira descreve o panorama da produção acadêmica, destacando as principais características dessa literatura em conjunto, fornecendo um mapa da temática discriminação algorítmica na pesquisa jurídica nacional. Já a segunda parte aprofunda a interpretação sobre os temas que estão em destaque no campo jurídico brasileiro relacionados ao assunto em questão. Por último, a seção de conclusões apresenta uma síntese da pesquisa realizada, juntamente com um prognóstico sobre a agenda futura para o aprimoramento dos estudos e do debate.

1 METODOLOGIA: REVISÃO DE ESCOPO APLICADA À PESQUISA

A técnica de revisão de literatura adotada nesta pesquisa seguiu a proposta do *Joanna Briggs Institute Reviewers' Manual* (PETERS *et al.*, 2015). As revisões de escopo têm a capacidade de mapear um tema de pesquisa, com o objetivo de compreender os conceitos-chave, as principais evidências, bem como as definições e limitações relacionadas ao assunto em questão (PETERS *et al.*, 2015). Portanto, nesta investigação, foi definido um formulário para análise das publicações sobre discriminação algorítmica com base em cinco informações, com apresentado na Tabela 1.

Tabela 1 – Formulário de análise para registro das informações-chave da revisão de escopo

Informação analisada	Descrição da coleta
----------------------	---------------------

RDP, Brasília, Vol.21 n. 110, 258-289, abr./jun. 2024, DOI: 10.11117/rdp.v21i110.7295 | ISSN:2236-1766



Tipo de discriminação	Extração dos casos discriminatórios citados em cada publicação
Ano da publicação	Coleta realizada a partir dos metadados das publicações
Áreas de pesquisa	Classificação dos artigos em áreas de pesquisa do Direito
Objeto de análise	Extração do objeto central da discussão das publicações analisadas
Problematização das análises	Mapeamento das causas e efeitos da discriminação algorítmica apresentadas pelas publicações

Fonte: Elaboração própria.

Dentre as informações registradas no formulário de análise da literatura, apenas o ano de publicação foi extraído dos metadados. As demais informações foram coletadas por meio da análise de conteúdo dos materiais selecionados para a revisão. Utilizou-se uma técnica exploratória de dados qualitativos para análise textual, com o objetivo de descrever e comparar o conteúdo dos textos (CAMARGO; JUSTO, 2013). Foram analisados os títulos, resumos e palavras-chave das publicações utilizando programação em *R*, onde foram executadas a mineração e análise textual. Nesse contexto, elaborou-se uma nuvem de palavras, que agrupa e organiza os termos presentes no texto graficamente em função de sua frequência, possibilitando uma visualização rápida dos principais termos-chave encontrados no conjunto textual (CAMARGO; JUSTO, 2013).

Os critérios estabelecidos para a inclusão e exclusão dos estudos selecionados são etapas fundamentais para garantir critérios confiáveis, rigorosos e passíveis de revisão e reprodução científica em revisões da literatura, visando reduzir subjetividades e garantir transparência nas escolhas feitas em relação ao que deve ser lido e relatado em um artigo ou seção de revisão (PETERS *et al.*, 2015). Para isso, foram realizadas buscas sistemáticas em três bases de dados amplamente reconhecidas no meio acadêmico para indexação de periódicos científicos: *Scopus*, *WoS* e *SciELO*.

A *Scopus* é atualmente a mais reconhecida fonte de dados de citação ao redor o mundo (SCHOTTEN *et al.*, 2018). A *WoS* tornou-se uma ferramenta internacional padrão para avaliação de pesquisas acadêmicas ainda na década de 1990 (SCHNELL, 2018). Já a *SciELO*,

há 25 anos, desempenha um papel fundamental como ferramenta de indexação e complemento às principais bases internacionais no Brasil. Além disso, essas três bases em conjunto constituem uma métrica importante para a classificação em quase todas as áreas da pós-graduação brasileira (MUGNAINI; NOYONS; PACKER, 2018).

Essas espécies de buscas sistemáticas são operacionalizadas por meio das chamadas *strings* de busca, que consistem nos termos inseridos no mecanismo de busca e funcionam com base em lógica matemática, permitindo a inclusão de prioridades e delimitações por meio de símbolos ou termos. Na Tabela 2 são apresentadas as *strings* utilizadas nas três bases de dados, juntamente com os resultados obtidos.

Tabela 2 – *Strings* de busca aplicadas nas bases *Scopus*, *Web of Science* e *SciELO*

Base	<i>String</i> de busca	Número de artigos coletados
<i>Scopus</i>	(TITLE-ABS-KEY ("Law" OR "Legal" OR "Juridical" OR "Judicial") AND TITLE-ABS-KEY ("Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Networks") AND TITLE-ABS-KEY ("Discrimination" OR "Bias" OR "Racism" OR "Algorithms" OR "Algorithmic Bias" OR "Algorithmic Discrimination" OR "Algorithmic Racism")) AND (LIMIT-TO (SUBJAREA , "SOCI")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (AFFILCOUNTRY , "Brazil"))	16
<i>Web of Science</i>	"Law" OR "Legal" OR "Juridical" OR "Judicial" (Topic) and "Artificial Intelligence" OR "Machine Learning" OR "Deep Learning" OR "Neural Networks" (Topic) and "Discrimination" OR "Bias" OR "Racism" OR "Algorithms" OR "Algorithmic Bias" OR "Algorithmic Discrimination" OR "Algorithmic Racism" (Topic) and Article (Document Types) and Law (Web of Science Categories) and BRAZIL (Countries/Regions)	8
<i>SciELO</i> ⁴		3

Obs. Buscas realizadas em 02 de maio de 2023.

Fonte: Elaboração própria.

⁴ Destaca-se que a *Web of Science* e a *SciELO*, que compartilham o mesmo mecanismo de busca. Essa integração ocorreu a partir do ano de 2014, quando a *SciELO* foi incorporada à plataforma da WoS.

Inicialmente, constata-se a existência de um número reduzido de publicações sobre o assunto no Brasil a partir dessas bases de dados, principalmente no recorte de periódicos indexados na área do Direito.⁵ Reconhece-se que o número de periódicos em Ciências Humanas e Sociais Aplicadas no Brasil que utilizam essas bases como indexadores não é significativamente alto. Para mitigar esse viés das bases (DONATO; DONATO, 2019), realizou-se a inclusão da chamada literatura cinza, que fornece contribuições relevantes para aprimorar as buscas sistemáticas, ampliando a sensibilidade e abrangência do procedimento de busca da literatura e mantendo-o gerenciável (PAEZ, 2017).

A inclusão dessa literatura foi realizada a partir de duas fontes de publicações. A primeira foi por meio do recurso de pesquisa avançada do GS, limitando-o ao Brasil e usando termos semelhantes aos utilizados nas bases anteriores.⁶ Isso resultou em um acréscimo de mais 17 artigos. Para obter uma amostragem mais representativa da produção acadêmica em Direito, a segunda fonte de inclusão de produtos acadêmicos foram os *papers* publicados nos *e-books* do CIDIA nas edições de 2020, 2021 e 2022. Esse último procedimento resultou na inclusão de mais 35 publicações.

A etapa inicial de identificação das publicações eliminou 5 artigos duplicados entre as bases *Scopus*, *WoS* e *SciELO*. No GS, foi criado o critério de exclusão com base no Qualis CAPES (2017-2020), eliminando os artigos publicados fora do estrato superior (A), resultando na exclusão de 8 artigos.⁷ Para a seleção das publicações do CIDIA, foram considerados os *papers* de autores vinculados a programas de pós-graduação, ou seja, aqueles assinados em autoria ou coautoria de mestrandos, mestres, doutorandos ou doutores, o que resultou na exclusão de 12 trabalhos. Na etapa de seleção das publicações para análise, foram realizadas leituras integrais dos textos e selecionados apenas aqueles que apresentavam discussões relevantes sobre a discriminação algorítmica, excluindo os textos que continham apenas citações protocolares do fenômeno ou que apenas o reconheciam. Por fim, esses processos de

⁵ Vale ressaltar que a *Scopus* não conta com um filtro exclusivo para a área do Direito, sendo essa parte das sociais aplicadas conforme está descrito na *string* (SUBJAREA , "SOCI"). Já na *WoS* e *SciELO* é possível usar um limitador específico para o campo do Direito, conforme explicado na busca "Law (Web of Science Categories)".

⁶ Os termos incluídos na busca do Google Scholar foram: "Lei" OR "Jurídico" OR "Judicial" OR "Jurídica" (Tópico) AND "Inteligência Artificial" OR "Aprendizado de Máquina" OR "Aprendizado Profundo" OR "Redes Neurais" AND "Discriminação" OR "Viés" OR "Racismo" OR "Algoritmos" OR "Viés Algorítmico" OR "Discriminação Algorítmica" OR "Racismo Algorítmico".

⁷ O estrato superior da CAPES é composto por periódicos classificados como A1, A2, A3 e A4 no último quadriênio (2017-2020).

A análise de tópicos revelou que a discriminação algorítmica, no âmbito da pesquisa jurídica, está sendo consolidada em pelo menos três eixos temáticos do Direito brasileiro, os quais se relacionam entre si. O primeiro eixo diz respeito aos impactos da IA no sistema judicial, com destaque para decisões automatizadas. O segundo eixo está relacionado às questões de proteção de direitos humanos e direitos fundamentais. Por fim, o terceiro eixo trata dos aspectos de transparência, ética e regulamentação da IA no Brasil.

Os impactos da IA no sistema de justiça brasileiro atravessam quase a totalidade das publicações analisadas. Mesmo que o objeto de análise do artigo investigado esteja afastado da problemática relacionada ao processo de decisão automatizada e dos reflexos dessas transformações no sistema judicial brasileiro, os efeitos da IA permeiam os eixos identificados na pesquisa. Contudo, os termos como “decisões”, “automatizadas” e “processo”, que possuem destaque na análise, exemplificam e representam esse eixo que tem foco em compreender os impactos e reflexos de julgamentos e decisões por meio de algoritmos de IA.

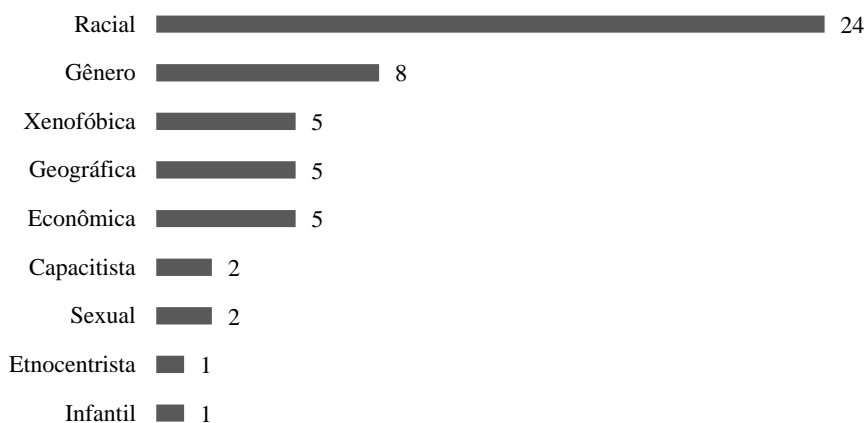
Entre os estudos analisados, aqueles voltados para compreender possíveis atentados contra direitos humanos ou violações de direitos fundamentais fornecem a maior aproximação com o fenômeno da discriminação algorítmica, problemática central desta revisão. Termos como “humanos”, “sociais” e “fundamentais” auxiliam no panorama desse eixo. Porém, vale destacar as palavras “racial” e “gênero” na nuvem, que serve como demarcadores de tipos de discriminação algorítmica com maior presença nos metadados dos textos investigados.

Por último, o eixo que envolve transparência de dados, opacidade algorítmica, regulamentação e ética na implementação da IA também se destaca de forma onipresente nas publicações que abordam a discriminação algorítmica. Além das discussões relacionadas à regulamentação das IAs no Brasil, como leis e resoluções que abordam o uso dessas tecnologias, os aspectos propositivos para a busca de soluções futuras ganham centralidade na atual conjuntura das discussões acadêmicas. Termos como “dados”, “explicabilidade”, “transparência”, “regulamentação” e “revisão” configuram-se na nuvem de palavras como temas centrais nessa abordagem na literatura.

Para aprofundar a descrição dos estudos sobre discriminação algorítmica, o Gráfico 1 apresenta os tipos de discriminações mencionados nas pesquisas analisadas, tendo como objetivo quantificar a frequência com que os tipos de discriminação por algoritmos são

mencionados e fornecer uma medida do acompanhamento realizado pela literatura publicada no Brasil.

Gráfico 1 – Tipos de discriminação algorítmica citados nas publicações sobre o tema no Brasil a partir da busca sistemática



Fonte: Elaboração própria.

A utilização de exemplos de discriminação está presente em 75% das publicações examinadas. O maior destaque é a predominância dos casos de racismo citados nas pesquisas sobre discriminação algorítmica, onde cerca de 69% dos estudos analisados mencionam pelo menos um caso de racismo algorítmico para exemplificar ou problematizar o fenômeno. Um dos casos mais emblemáticos e amplamente citados entre as pesquisas analisadas é o caso do COMPAS (*Correctional Offender Management profiling for Alternative Sanctions*) nos Estados Unidos. Esse algoritmo, utilizado no sistema de processo penal estadunidense, discriminava réus negros ao classificá-los com maior probabilidade de cometer crimes futuros em comparação com pessoas brancas (ANGWIN *et al.*, 2016). Casos relacionados ao reconhecimento facial também são frequentemente mencionados como exemplos de racismo, como o incidente em que o Google rotulou erroneamente a imagem de dois jovens negros como gorilas (BBC, 2015).

A discriminação de gênero foi a segunda mais frequente entre as publicações. Um caso bastante conhecido de discriminação de gênero promovida por algoritmos ocorreu no âmbito online com a *chatbot* Tay, desenvolvida pela Microsoft. Por meio da interação com usuários do Twitter, a robô adquiriu comportamentos fortemente misóginos e transfóbicos, além de

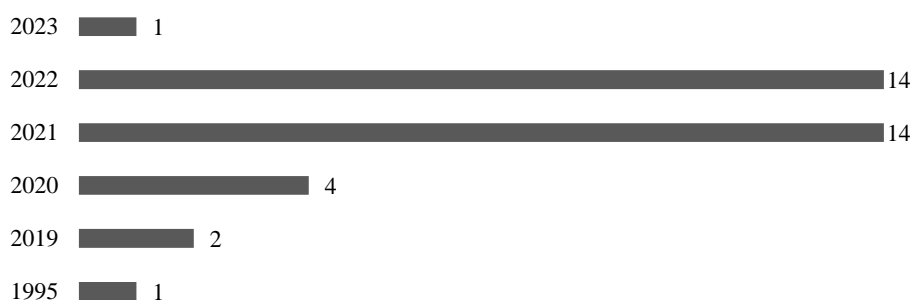
xenófobos (VINCENT, 2016). Além disso, casos relacionados a vagas de emprego também são citados, como o algoritmo de recrutamento de candidatos da Amazon, que, devido à predominância de currículos masculinos, passou a reduzir a pontuação de mulheres, diminuindo suas chances de contratação (DASTIN, 2018). Também houve relatos de discriminação sexual contra grupos LGBTQIA+ no controle de conteúdo de plataformas de redes sociais, em que termos utilizados na comunicação desses grupos foram classificados como tóxicos por algoritmos de análise automatizada de conteúdo (OLIVEIRA, 2020).

No geral, os casos de discriminação apresentam intersecções significativas. No entanto, as discriminações econômicas e geográficas estão mais associadas nos exemplos citados. Isso ocorre porque parte dos casos mencionados nos estudos envolviam características em que a localização dos indivíduos indicava condições econômicas menos favoráveis. Essa prática, conhecida como *geopricing*, ganhou notoriedade no Brasil com o caso da empresa Decolar, que aplicava diferentes preços de acordo com a localização do consumidor, além de impossibilitar reservas com base na geolocalização do usuário (VEJA, 2018). Os casos de discriminação socioeconômica vão além de exemplos materiais e também são mencionados em relação às condições econômicas e sociais do país, como é o caso do Brasil. Por exemplo, níveis de desigualdades educacionais podem aprofundar diferenças nos processos de seleção para vagas de emprego ou até mesmo acesso à universidade, assim como na concessão de créditos e financiamentos (*credit score*), que estão associados à localização do consumidor, limitando as oportunidades daqueles que residem em áreas de baixa renda (FRANÇA NETTO; EHRHARDT JÚNIOR, 2022; HENRIQUES; SAMPAIO, 2021).

Outros tipos de discriminação, como o capacitismo, foram relatados em menor quantidade nas publicações, mas destacam os riscos da discriminação algorítmica em relação a pessoas que sofrem de alguma limitação física ou mental. Por exemplo, há o caso de um jovem dos EUA que, após ser diagnosticado com transtorno bipolar, enfrentou dificuldades para encontrar emprego devido a um teste de personalidade desenvolvido pela empresa Kronos, que incluía perguntas sobre saúde mental (O'NEIL, 2018). Adicionalmente, é importante mencionar a discriminação cultural e etnocêntrica, que se refere aos padrões ocidentais das diretrizes de programação adotados por grandes conglomerados empresariais do setor. Esses padrões estão sujeitos a falhas ao serem aplicados em contextos regionais específicos, com diversidades sociais, culturais e jurídicas (ADAM; FURNIVAL, 1995).

A observação seguinte refere-se ao desenvolvimento temporal das publicações sobre a temática. A análise no Gráfico 2, demonstra a recente popularização da discussão e como o processo de democratização de acesso às ferramentas de IA têm sido absorvidos pela produção acadêmica no Direito.

Gráfico 2 – Ano das publicações sobre discriminação algorítmica no Brasil a partir da busca sistemática



Fonte: Elaboração própria.

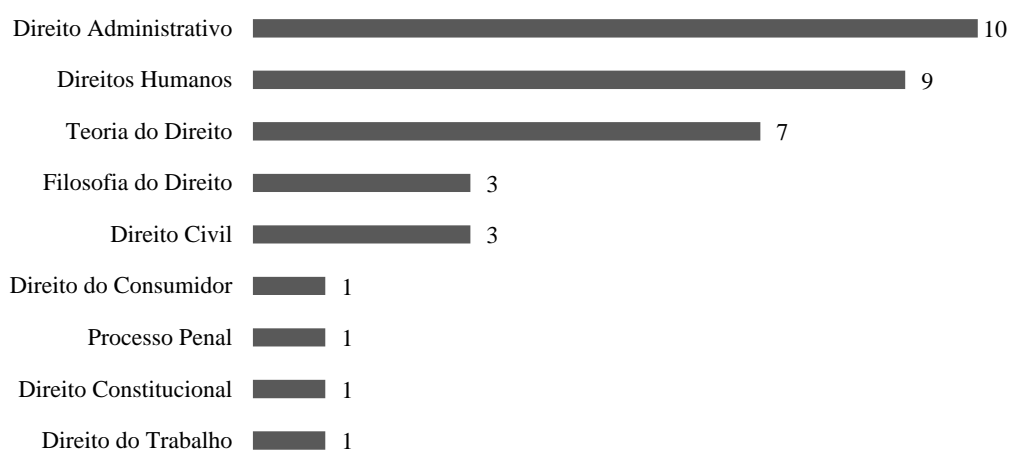
Entre os anos em que foram publicados os textos analisados pela revisão proposta, destaca-se o ano de 1995, em total contraste com o padrão da análise que se concentra nos últimos cinco anos. Especificamente, foi identificado um artigo publicado no periódico *Information & communication technology law*. Neste artigo, as autoras Alison Adam e Chloë Furnival (ADAM; FURNIVAL, 1995), esta última professora do Departamento de Ciência da Informação da Universidade Federal de São Carlos, analisam o sistema de inteligência artificial *Cyc* sob a perspectiva da teoria jurídica feminista. As autoras questionam a reprodução de estruturas de discriminação e desigualdades existentes por parte desse sistema de IA.

A partir de 2019, é possível observar um crescimento significativo no número de publicações que abordam o tema da discriminação algorítmica. Nos anos de 2021 e 2022, em particular, foram analisadas 14 publicações em cada um desses anos. Essa tendência ascendente indica um aumento de interesse e engajamento dos pesquisadores com essa temática. Um olhar para as bases de dados utilizadas revelou que incentivos institucionais, como dossiês em periódicos, promovem um papel importante na promoção e estímulo aos pesquisadores a explorarem essas temáticas. Por exemplo, a Revista de Direito Público publicou o dossiê “Inteligência artificial, ética e epistemologia” em 2021 (WIMMER; DONEDA, 2021). Além

disso, o Congresso Internacional de Direito e Inteligência Artificial realizou três edições consecutivas nos anos de 2020, 2021 e 2022.

O mapeamento a seguir descreve as áreas do Direito brasileiro que abordam a temática da discriminação algorítmica. A partir da amostragem de publicações que exploram essa problemática, foi possível identificar quais áreas de pesquisa das ciências jurídicas estão se dedicando à compreensão desse fenômeno. No Gráfico 3, são apresentadas as áreas identificadas de cada uma das publicações analisadas.

Gráfico 3 – Subáreas de pesquisa das publicações sobre discriminação algorítmica no Brasil a partir da busca sistemática



Fonte: Elaboração própria.

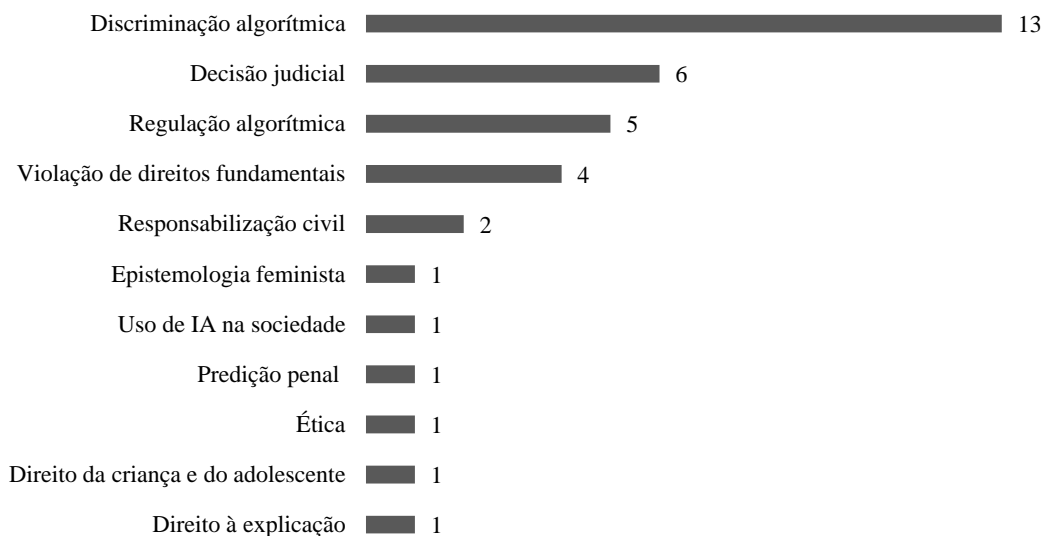
Os resultados da categorização das publicações por área de pesquisa no Direito corroboram com a exploração apresentada na Figura 1, que exibe a nuvem de palavras a partir dos metadados dos autores. Observa-se um predomínio das áreas do Direito Administrativo e dos Direitos Humanos nas pesquisas que tratam a temática da discriminação algorítmica, seguidas pelos estudos da Teoria do Direito e Direito Civil, que estão principalmente relacionados ao primeiro eixo mencionado na análise anterior. Esse eixo enfoca os impactos e efeitos da utilização de algoritmos no sistema de justiça, abrangendo desde julgamentos e decisões automatizadas até o processo de responsabilização civil de danos causados por essas ferramentas.

Inserido nas publicações que permeiam o campo da Filosofia do Direito, identifica-se o estudo já citado de epistemologia feminista para análise do fenômeno tecnológico no campo jurídico. No entanto, vale ressaltar atenção dada ao papel da bioética e às mutações do ser humano, assim como possíveis práticas eugênicas de discriminação (TOBBIN; CARDIN, 2022). Já os demais ramos do Direito apresentam números significativamente inferiores de publicações que tocam o tema da discriminação algorítmica. Abordagens mais especializadas, como o Direito do Trabalho ou o Direito do Consumidor, possuem publicações em menor número nessa amostragem, onde destaca-se o caso do papel dos sindicatos frente à violação de direitos fundamentais dos trabalhadores (MARQUES; NETO MARTINEZ, 2022) ou da discriminação algorítmica de consumidores com base em suas geolocalizações (ALVES; ROCHA, 2021).

Devido à contemporaneidade e relevância do tema na sociedade, é razoável concluir que o campo do Direito deve aprofundar cada vez mais os estudos sobre os impactos dos algoritmos, tanto no sistema de justiça quanto na sociedade em geral. Contudo, os campos de pesquisa que assumem baixa concentração de publicações são indicativos de lacunas e espaços a serem preenchidos por pesquisadores nos próximos anos.

Para prosseguir na exploração dos resultados, no Gráfico 4 são exibidos os objetos de análise, ou seja, o principal assunto abordado pelos autores em suas pesquisas.

Gráfico 4 – Objetos de análise das publicações sobre discriminação algorítmica no Brasil a partir da busca sistemática



Fonte: Elaboração própria.

Constata-se que 36% dos textos analisados têm como foco a discriminação algorítmica, o que significa dizer que 64% dos artigos examinados nesta revisão contemplam a discriminação algorítmica como um objeto correlato em suas discussões. As pesquisas com foco nas decisões judiciais e na regulamentação de algoritmos de IA receberam maior atenção nos estudos acadêmicos dentro no recorte analisado.

É possível observar que, de forma geral, os objetos de pesquisa se entrecruzam, mas podem ser investigados em maior ou menor escala por campos distintos das Ciências Jurídicas, como é principalmente o caso da discriminação algorítmica e dos Direitos Humanos. Em contrapartida, identifica-se que objetos de pesquisa como a regulamentação de algoritmos de IA se concentram no ramo do Direito Administrativo, assim como os estudos sobre decisões judiciais estão sendo abordados principalmente pela Teoria do Direito.

As análises dos objetos centrais dessas investigações auxiliaram a revelar as principais problematizações e preocupações abordadas por esses estudos quando mobilizam o conceito de discriminação algorítmica. Os fatores discriminatórios e o aprofundamento das desigualdades sociais, juntamente com o enfoque no problema da transparência e regulamentação dos algoritmos constituem os principais pontos de atenção dos pesquisadores da área em relação à discriminação algorítmica.

Particularmente, a questão da transposição das desigualdades do mundo real para o digital no Brasil vem sendo estudada pela doutrina jurídica recente (GABARDO; VIANA; FREITAS, 2022). O fato de a discriminação por meio de algoritmos reproduzir discriminações sociais pré-existentes se mostra quase um consenso entre os investigadores do tema. Um exemplo de análise que centraliza as desigualdades sociais no contexto dos algoritmos é o artigo de França Netto e Ehrhardt Júnior (2022), que discute os impactos da discriminação algorítmica no cenário brasileiro, uma vez que essas discriminações recaem sobre grupos historicamente excluídos. Outro exemplo de referência que apresentou um panorama das desigualdades sociais junto ao avanço dessas ferramentas de IA foi o estudo de Henriques e Sampaio (2021) sobre a exclusão tecnológica de crianças, que são mais vulneráveis e estão em processo de formação, o que significa que sentirão os efeitos das novas tecnologias por mais tempo (HENRIQUES; SAMPAIO, 2021, p. 252).

A violação de direitos humanos é caracterizada nessa literatura como a violação de importantes princípios por meio dos recentes avanços tecnológicos e com regulamentações insuficientes. Trata-se de lesões aos direitos de personalidade, privacidade e honra, bem como abusos do Estado com ameaças autoritárias e de controle social sobre a vida (WERMUTH; CARDIN; WOLOWSKI, 2021). A falta de aderência aos direitos humanos internacionais por parte de sites de redes sociais também é alvo das pesquisas, uma vez que a moderação de conteúdo em plataforma de redes sociais sem a fiscalização humana acaba suprimindo os direitos dos usuários e a liberdade de expressão, o que já foi constatado em relação a grupos minoritários e historicamente excluídos, como a comunidade LGBTQIA+ e movimentos antirracistas (OLIVEIRA, 2020). Grupos estes que apesar de estarem obtendo vitórias no âmbito jurisprudencial brasileiro, estão vivenciando uma ampliação dos ataques neoconservadores da extrema-direita, notadamente no âmbito das redes sociais (RIOS, 2022).

Os pesquisadores que direcionaram seus esforços para compreender o problema da transparência e regulamentação dos algoritmos conectam a problematização dos vieses discriminatórios com a opacidade dos algoritmos de IA. Segundo a literatura, o desconhecimento em relação aos *softwares* de IA, desde o tratamento de informações pessoais (dados sensíveis) até a etapa de programação da IA, é responsável por potencializar a geração de *outputs* discriminatórios (ALVES; ANDRADE, 2022; ARAÚJO; ZULLO; TORRES, 2020; MENDES; MATTIUZZO, 2019; TONIAZZO; BARBOSA; RUARO, 2021). Além disso, esses

estudos abordam a regulamentação no Brasil ou em perspectiva comparada, citando dispositivos legais (leis e resoluções) em vigor ou em tramitação no Legislativo brasileiro (CAVALCANTE, 2022; GOMES; VAZ; DIAS, 2021).

Os demais tópicos identificados pela análise revelam preocupações subjacentes, apontando tendências e lacunas específicas. A baixa visibilidade de problemas como responsabilidade civil, *XAI* e reprodução de desigualdades sociais pode ser explicada pela complexidade técnica e volatilidade do assunto. O rápido avanço dessas tecnologias dificulta a aproximação de pesquisadores de áreas correlatas, tornando essas questões emergentes no campo de investigação. Essas constatações indicam que, cada vez mais, é necessária uma compreensão interdisciplinar para aperfeiçoar o desenvolvimento dessas tecnologias em relação aos problemas jurídicos e sociais futuros.

A compreensão sobre a responsabilidade civil em casos de discriminação causada por IA demonstra uma dessas lacunas presentes na literatura. Ainda que em pequeno número, os estudos identificados com essa preocupação argumentam que, no atual ordenamento jurídico brasileiro, a falta de personalidade jurídica da IA impede sua responsabilização civil em casos de discriminação. Isso tem sido utilizado por empresas desenvolvedoras de algoritmos como uma forma de evitar reparações por danos causados pela IA (COSTA, 2023). O exemplo citado por Medeiros e Moção (2021) ilustra o dilema da responsabilização civil em processos automatizados ao questionar: “de quem é a responsabilidade pela decisão assistida de não contratar determinado candidato, a empresa ou o agente desenvolvedor do sistema de IA?”. Sob essa perspectiva, esses autores ressaltam a importância de práticas de *accountability* no desenvolvimento dos algoritmos, incluindo responsabilidade ética, transparência e prestação de contas, além da implementação de medidas preventivas contra a discriminação algorítmica (COSTA, 2023; MEDEIROS; MOÇÃO, 2021).

Entre as publicações examinadas, uma análise abordou a preocupação com o impacto da IA no processo penal. Nessa publicação, a discriminação algorítmica foi discutida de forma paralela ao tema principal, destacando os resultados preditivos no processo penal provenientes de tecnologias alimentadas por seres humanos, ou seja, tecnologias não neutras (GUIMARÃES, 2019). O debate sobre políticas públicas relacionadas a esses fenômenos foi identificado apenas uma vez, como uma possibilidade para mitigar a discriminação algorítmica. Essa estratégia propõe a adoção de ações afirmativas que envolvam a inclusão de pessoas pertencentes a grupos

minoritários nas equipes de desenvolvimento dessas tecnologias, visando trazer diversidade social para o campo profissional da programação (REQUIÃO; COSTA, 2022). Além disso, nas análises, o aspecto da desumanização do trabalho se manifestou por meio dos riscos da vigilância algorítmica sobre os trabalhadores. Por exemplo, medidas de desempenho adotadas durante a jornada de trabalho podem resultar em graves violações de intimidade e privacidade, além de promover discriminação social por meio de comparações físicas e tornar os trabalhadores suscetíveis a doenças psicológicas (MARQUES; NETO MARTINEZ, 2022).

2.2 A DISCRIMINAÇÃO ALGORÍTMICA NA ÓPTICA DOS DIREITOS HUMANOS E FUNDAMENTAIS

Nesta seção do artigo empreende-se um esforço para proporcionar uma interpretação do conhecimento consolidado na temática em análise. Trata-se de uma abordagem referencial do debate estabelecido por pesquisadores da área do Direito no Brasil que investigam a discriminação algorítmica à luz dos direitos humanos e fundamentais. O primeiro passo para isso é compreender que a violação de direitos por algoritmos de IA pode ocorrer de várias formas, algumas das quais são mais objetivas, como casos de discriminação direta ou indireta contra minorias sociais (FRANÇA NETTO; EHRHARDT JÚNIOR, 2022). Além disso, também pode haver violação de princípios processuais (FORTES, 2020) e direitos como liberdade de expressão, privacidade ou personalidade (MARQUES; NETO MARTINEZ, 2022; OLIVEIRA, 2020; TOBBIN; CARDIN, 2022).

Conforme explicado por Duarte e Negócio (2021), a reflexão sobre discriminação algorítmica tem como base o princípio da igualdade. Além do princípio fundamental da igualdade, devem ser considerados os direitos à diversidade e à diferença, o que remete ao direito fundamental à não-discriminação, princípio previsto em convenções globais de Direitos Humanos (ONU, 1948, 1966). Portanto, a igualdade é interpretada por meio de uma dimensão substancial, que abrange conceitos como justiça social e distributiva, e “se reflete na busca pela inclusão social daqueles indivíduos excluídos e subordinados devido a características como cor, gênero, orientação sexual, entre outras.” (DUARTE; NEGÓCIO, 2021, p. 6). Ademais, como direito fundamental, possui tanto uma dimensão objetiva quanto uma dimensão subjetiva – o

que implica o reconhecimento de pretensões jurídicas jusfundamentais específicas (HACHEM, 2019).

Nesse sentido, reforça-se o entendimento de que as discriminações manifestadas por algoritmos têm o caráter de reproduzir as desigualdades e discriminações sociais pré-existentes na estrutura social, resultando na manutenção de grupos minoritários em posições socialmente subordinadas (PINCUS, 2018). A intencionalidade discriminatória pode estar presente na programação dessas ferramentas, como na categorização de dados, na atribuição de pesos estatísticos e no enviesamento de bancos de dados (FRANÇA NETTO; EHRHARDT JÚNIOR, 2022). No entanto, também podem ocorrer em um cenário aparentemente neutro, desprovido de vieses, negligenciando as diferenças sociais desde a marginalização e exclusão de grupos sociais nos *Big Data*, até preconceitos ocultos por meio de correlações e codificações redundantes (BAROCAS; SELBST, 2016; DUARTE; NEGÓCIO, 2021). As decisões humanas sempre foram influenciadas por enviesamentos. A utilização de algoritmos pode amplificar esses preconceitos sociais, aumentando o alcance das decisões enviesadas pelas máquinas. (PUSCHEL; RODRIGUES; VALLE, 2022; RIBEIRO, 2022). Um exemplo evidente disso é encontrado no estudo de Fortes (2020, p. 5) no qual é discutido o enviesamento estatístico do caso COMPAS, que resultou em atribuições de pontuações de risco desproporcionais entre réus negros e brancos.

A intensificação das possíveis violações de direitos humanos e fundamentais por parte dos algoritmos de IA também afeta o direito à liberdade de expressão. Embora a mobilização das tecnologias possam potenciais de melhorar a qualidade do conteúdo online, combatendo, por exemplo, discursos de ódio e manifestações preconceituosas (o que é uma exigência contemporânea), há relatos de bloqueio de conteúdo que privam certos grupos minoritários de se expressarem livremente devido a aspectos linguísticos e culturais, contrariando princípios estabelecidos em tratados internacionais (ONU, 1966; OLIVEIRA, 2020). Um exemplo disso são as redes sociais que utilizam ferramentas algorítmicas para moderar o conteúdo, às vezes sem controle humano ou com uma revisão tardia, não atendendo aos requisitos estabelecidos pelo *International Covenant on Civil and Political Rights* (OLIVEIRA, 2020).

O avanço das novas tecnologias transforma os direitos fundamentais de terceira dimensão por meio do controle da liberdade e intimidade dos indivíduos, devido ao compartilhamento diário de múltiplos dados em rede (SARLET, 2018, 2021). Diante de

violações constantes, os direitos da personalidade também são objeto de debates e preocupações nesses estudos. A violação dos direitos à imagem, honra e intimidade dos indivíduos por meio da exposição de dados pessoais tem se tornado cada vez mais comum (WERMUTH; CARDIN; WOLOWSKI, 2021). Os riscos decorrentes do avanço do controle social com base em informações pessoais, que podem criar condições para políticas totalitárias respaldadas juridicamente, capazes de controlar a vida e a morte dos indivíduos, são motivo de preocupação para os pesquisadores (WERMUTH; CARDIN; WOLOWSKI, 2021).

Da mesma forma, Henriques e Sampaio (2021, p. 261) argumentam que “o uso indiscriminado de dados pessoais é capaz de objetificar as pessoas, promover manipulações, afetar o livre desenvolvimento da personalidade e gerar discriminações”. Essa afirmação está relacionada, por exemplo, a violações dos direitos fundamentais trabalhistas que afetam os direitos à privacidade, intimidade e dignidade humana, como nos casos de *softwares* de monitoramento de desempenho e comportamento dos trabalhadores, que podem ter potencial extremamente invasivo, resultando em excesso de autocobrança e posterior adoecimento dos indivíduos (MARQUES; NETO MARTINEZ, 2022).

A exposição dessa discussão evidencia o papel das instituições perante o desenvolvimento responsável da IA. Assim, é imperativo que as instituições e seus agentes atuem com transparência e responsabilidade, assegurando que os avanços tecnológicos promovam mudanças positivas e justas na sociedade, contribuam para o bem-estar social e respeitem os direitos fundamentais consolidados.

2.3 A TRANSPARÊNCIA E REGULAMENTAÇÃO E A DISCUSSÃO SOBRE FORMAS DE MITIGAR A DISCRIMINAÇÃO ALGORÍTMICA

Diante dos potenciais riscos dos avanços tecnológicos em relação aos direitos fundamentais, especialmente aqueles relacionados aos dados pessoais, diversas organizações internacionais têm se engajado nas últimas décadas em discussões sobre diretrizes e regulamentos para a utilização responsável e ética dessas ferramentas, (EUROPEAN UNION, 2016; OECD, 2019) por meio de princípios e recomendações pautadas em uma IA baseada em valores humanos (BRAVO, 2020). Nesse sentido, o Brasil também adotou medidas para regulamentar e estabelecer diretrizes que visam mitigar e combater possíveis discriminações e

abusos de poder decorrentes de bancos de dados e do uso de algoritmos de IA, destacando-se a Lei nº 13.709/18 – Geral de Proteção de Dados (LGPD) e, ainda, a Portaria nº 4.617/2021, que estabelece a Estratégia Brasileira de Inteligência Artificial (EBIA) sob responsabilidade do Ministério da Ciência, Tecnologia e Inovação.

A EBIA fundamenta-se nas diretrizes estabelecidas pela *Organisation for Economic Co-operation and Development* (OECD), da qual o Brasil é signatário. A Portaria nº 4.617/2021 do MCT&I destaca alguns princípios levantados por essa organização internacional, tais como: (I) os benefícios da IA devem ser centrados no ser humano, no planeta, no desenvolvimento sustentável e no bem-estar; (II) o dever de respeitar o Estado Democrático de Direito, os Direitos Humanos e a diversidade social; (III) a transparência e a divulgação responsável das ferramentas de IA, permitindo explicação, revisão e contestação por parte dos usuários e das pessoas afetadas; e (IV) o funcionamento robusto, seguro e protegido, com avaliações contínuas.

No entanto, a doutrina jurídica brasileira concentra grande parte de seus esforços para discutir os riscos e limites da aplicação dos algoritmos em diversas áreas da sociedade brasileira, especialmente levando-se em consideração a multifuncionalidade dos direitos fundamentais na Lei Geral de Proteção de Dados – LGPD (HIROMI; SALGADO, 2020). O que se tem destacado é uma preocupação significativa relacionada ao uso de IA no sistema judicial brasileiro, devido ao potencial discriminatório que essas ferramentas podem apresentar em decisões automatizadas. A partir de 2020, o Conselho Nacional de Justiça estabeleceu diretrizes para o uso da IA em conformidade com a LGPD e com os artigos 195 e 196 do Código de Processo Civil, fundamentando-se no princípio da eficácia (CNJ, 2020a, b). Os estudos sobre esse tema abordam o direito à explicação previsto na LGPD em casos de tratamento de dados pessoais por meio de decisões automatizadas, questionando a capacidade da legislação em lidar com a opacidade algorítmica (TONIAZZO; BARBOSA; RUARO, 2021).

Com base nessa abordagem, Toniazzo, Barbosa e Ruaro (2021) discutem as limitações gerais previstas no ordenamento jurídico brasileiro em relação ao direito à explicação. O princípio da explicabilidade é considerado fundamental para estabelecer relações de confiança nos sistemas de IA tornando os processos transparentes e buscando alcançar decisões imparciais e com menor grau de subjetividade.

Existe uma interpretação restritiva e equivocada da LGPD que sugere que o direito à revisão por parte de uma pessoa natural não está previsto em seu art. 20. Parece mais adequada ao sistema constitucional brasileiro, todavia, a realização de uma hermenêutica extensiva para garantir o direito à explicação e os interesses dos titulares de dados pessoais (TONIAZZO; BARBOSA; RUARO, 2021). No entanto, há também o argumento de que o art. 20 da Lei nº 13.709/2018 reconhece o direito “de solicitar a revisão de uma decisão automatizada por uma pessoa natural, caso o titular dos dados considere que seus interesses foram afetados” (GOMES; VAZ; DIAS, 2021, p. 116).⁸ Nessa perspectiva, o direito à explicação e o princípio da transparência desempenham um papel fundamental no avanço do debate em direção a uma inteligência artificial ética e fundamentada nos direitos humanos, como apontam Cardoso e Pessoa (2022, p. 93).

Para promover uma inteligência artificial ética, a doutrina jurídica brasileira vem incorporando o conceito de “desenviesamento” humano e algorítmico, com o objetivo de aplicar o princípio da não-discriminação (RIBEIRO, 2022). Índícios dessa busca pelo combate aos vieses na utilização da IA podem ser observados na Resolução 332 do CNJ, que estabelece a criação de condições para eliminar ou minimizar decisões discriminatórias baseadas em preconceitos, além de exigir a homologação de modelos de IA capazes de “identificar se preconceitos ou generalizações influenciaram seu desenvolvimento” (CNJ, 2020b). No entanto, uma crítica relevante feita nessa literatura está relacionada à aparente omissão da regulamentação do CNJ em relação à origem dos dados que alimentam os algoritmos de IA, os quais provêm de uma sociedade marcada por preconceitos e desigualdades (CARDOSO; PESSOA, 2022). Essa questão ressalta a importância de considerar os agentes responsáveis pela construção dos algoritmos e pela implementação das IAs na sociedade. Não se pode negligenciar ou abstrair dessa discussão jurídica os interesses por trás dos tratamentos de bancos de dados e da programação dos algoritmos ao regulamentar esses processos, pois a responsabilidade desses atores é crucial para garantir uma IA que seja justa e equitativa.

Se o desenviesamento se torna um antídoto importante para combater a discriminação algorítmica e alcançar decisões éticas, transparentes e imparciais no uso da IA na sociedade (PUSCHEL; RODRIGUES; VALLE, 2022), o próprio avanço tecnológico passa a desempenhar

⁸ Vale ressaltar que o trecho da lei que previa a revisão de decisões automatizadas por uma pessoa natural foi retirado do texto em lei por veto presidencial.

um papel fundamental nesse progresso. Alves e Andrade (2022) descrevem os avanços em direção a uma Inteligência Artificial Explicável (*XAI*), na qual permite combater a opacidade algorítmica por meio de seu próprio sistema preditivo, fornecendo explicações sobre o funcionamento de suas correlações internas. A chamada "caixa de vidro" da *XAI* promove uma maior confiabilidade para os usuários e a sociedade em geral, permitindo a verificação e a atribuição de responsabilidade às decisões tomadas pela IA (ALVES; ANDRADE, 2022). Todavia, resta evidente que estas novas ferramentas de transparência possuem limites técnicos autorreferentes. O uso de redes neurais artificiais e *deep learning*, implicam o reconhecimento e aceitação de algum nível de inexplicabilidade, pela própria natureza probabilística e autônoma de funcionamento do sistema na atualidade. Esta é uma questão permanente e que exige acompanhamento, controle e regulação.

Por fim, há ainda a preocupação em relação às automatizações no âmbito da gestão pública e às possíveis limitações dos conceitos e medidas mais discutidos, como transparência administrativa e *accountability*. Essa preocupação é abordada por Mendes e Mattiuzzo (2019), que destacam as limitações dessas medidas, especialmente quando se trata de informações sensíveis e motivos privados. Logo, há a necessidade de combinar estratégias e considerar aspectos de responsabilidade pela Administração Pública na concepção dos modelos de IA, desde a formação das equipes de programadores até os últimos processos de execução dos serviços. Isso implica capacitar as pessoas não apenas em questões técnicas de programação, mas também em princípios éticos e de governança, a fim de compreenderem os impactos das tecnologias no mundo real (MENDES; MATTIUZZO, 2019).

CONSIDERAÇÕES FINAIS

Ao perseguir o objetivo de analisar como a discriminação algorítmica tem sido abordada na pesquisa acadêmica no campo do Direito no Brasil e, ainda, fornecer uma descrição do estado atual dessa discussão, a investigação revelou uma consolidação de dois eixos de análise distintos. O primeiro eixo está relacionado às abordagens fundamentadas nos direitos humanos e fundamentais, enquanto o segundo aborda o fenômeno sob a perspectiva da regulamentação e transparência dos algoritmos. Essa descoberta corrobora a hipótese formulada no problema de pesquisa e, na prática, implica não apenas na concentração de estudos em casos

discriminatórios semelhantes, mas também na identificação de objetos de análise e problemáticas similares, além de lacunas de investigação em outras áreas correlatas do Direito.

Evidenciou-se uma temática de pesquisa recente, com crescimento significativo nos últimos dois anos, período em que 28 dos 36 textos analisados foram publicados. Esse panorama sugere a crescente relevância dessa temática no campo acadêmico nacional, despertando interesse não apenas nas publicações em periódicos, mas também nas motivações pessoais dos pesquisadores, especialmente aqueles em nível de pós-graduação. Além disso, foi possível identificar o crescimento dessa temática em diversas subáreas do Direito, embora algumas delas já desponham com maior dedicação dos pesquisadores, como as abordagens no âmbito do Direito Civil e Direito do Trabalho.

Ao observar a importância da responsabilização civil na discussão sobre a atribuição de responsabilidade em casos de violações de direitos causadas por IA, verifica-se essa como um exemplo de lacuna do campo acadêmico. O mesmo pode ser observado em relação aos impactos desse fenômeno no contexto das relações de trabalho. Apesar dos esforços recentes das instituições judiciais para tornar o tema objeto de atenção, ainda é necessário um maior aprofundamento nas explicações sobre processos preditivos de contratação e controle sobre as violações de direitos à privacidade dos trabalhadores.

Destaca-se, ainda, que a temática da decisão judicial é significativamente abordada nas publicações, principalmente atreladas aos debates da Teoria do Direito. Contudo, percebe-se que o problema da decisão por uma perspectiva do Processo Penal não tem recebido atenção semelhante, evidenciando uma lacuna. Já um ponto consensual dessa bibliografia é a necessidade de regulação estatal do uso da inteligência artificial com o objetivo de evitar a discriminação, notadamente por intermédio da imposição do dever de maior transparência. Nesse sentido, o material pesquisado confirma proposições como a de Fábio Sousa Santos (2021a, p. 255) de que uma sociedade cada vez mais digital requer um aparato estatal forte e capaz de responder aos desafios impostos – contraditando, portanto, a visão liberal a respeito do fenômeno.

Especula-se que à medida que os sistemas de IA forem mais amplamente empregados no Brasil e o conhecimento sobre suas aplicações na sociedade se tornar mais difundido, haverá uma disseminação da produção acadêmica em seus diversos campos e subcampos de pesquisa. Ademais, as inovações deverão implicar uma regulação à luz da precaução e do

desenvolvimento sustentável (SCHIER; MAKSYM; MOTA, 2021). E é necessário ressaltar que, assim como em todo o campo de pesquisa dedicado à IA, o Direito também requer investigações interdisciplinares, a fim de avançar no desenvolvimento de diagnósticos atualizados e habilitados para elaborar adequadas políticas públicas e possíveis soluções regulatórias para as tecnologias na sociedade a partir dos comandos do novo “constitucionalismo digital” (SOUSA; CRISTÓVAM; MACHADO, 2022).

REFERÊNCIAS

ADAM, Alison; FURNIVAL, Chloe. Designing Intersections-Designing Subjectivity: Feminist Theory and Praxis in a Sex Discrimination Legislation System. **Information & Communications Technology Law**, v. 4, n. 2, 1995.

ALVES, Fabrício Germano; ROCHA, Vinícius Wdson do Vale. Discriminação na precificação conforme a geolocalização (geopricing) do consumidor. In: **II Congresso Internacional de Direito e Inteligência Artificial**, 2021, Belo Horizonte: [s.n.], 2021.

ALVES, Marco; ANDRADE, Otávio Morato de. Da “caixa-preta” à “caixa de vidro”: o uso da explainable artificial intelligence (XAI) para reduzir a opacidade e enfrentar o enviesamento em modelos algorítmicos. **Direito Público**, v. 18, n. 100, p. 349–373, 2022.

ANGWIN, Julia *et al.* **Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks**, 2016. Disponível em:

<<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>. Acesso em: 10 jun. 2023.

ARAÚJO, Adriane Reis de *et al.* **O uso de dados pessoais e inteligência artificial na relação de trabalho: proteção, discriminação, violência e assédio digital**. Brasília: Ministério Público do Trabalho, 2022.

ARAÚJO, Valter Shuenquener de; ZULLO, Bruno Almeida; TORRES, Maurílio. Big Data, algoritmos e inteligência artificial na Administração Pública: reflexões para a sua utilização em um ambiente democrático. **A&C – REVISTA DE DIREITO ADMINISTRATIVO & CONSTITUCIONAL**, v. 20, n. 80, p. 241–261, 2020.

BAROCAS, Solon; SELBST, Andrew D. Big Data ’ S Disparate Impact. **California Law Review**, v. 104, n. 3, p. 671–732, 2016.

BBC. **Google apologises for Photos app’s racist blunder**, 2015. Disponível em: <<https://www.bbc.com/news/technology-33347866>>. Acesso em: 10 jun. 2023.

BRASIL. Ministério da Ciência, Tecnologia e Inovações. **Portaria nº 4.617**, de 6 de abril de 2021. Diário Oficial da União, Edição 67, Seção 1, p. 30, 12 abr. 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-portaria_mcti_4-617_2021.pdf.



Acesso em: 20 jun. 2024.

BRAVO, Álvaro A. Sánchez. Marco europeo para uma inteligência artificial baseada em las personas. **International Journal of Digital Law**. v. 1, n. 1, p. 65-78, 2020.

CAMARGO, Brígido Vizeu; JUSTO, Ana Maria. IRAMUTEQ: Um Software Gratuito para Análise de Dados Textuais. **Temas em Psicologia**, v. 21, n. 2, p. 513-518, 2013.

CARDOSO, Henrique Ribeiro; PESSOA, Flávia Moreira Guimarães. Inteligência artificial e julgamento por computadores: uma análise sob a perspectiva de um acesso a justiça substancial. **Revista Jurídica**, v. 5, n. 72, p. 75-101, 2022.

CAVALCANTE, Jamile Sabbad Carecho. Um panorama do avanço legislativo acerca da regulamentação da inteligência artificial no Brasil. 2022, [S.l: s.n.], 2022.

CONSELHO NACIONAL DE JUSTIÇA. **Resolução nº 332**, de 21 de agosto de 2020. Disponível em: <<https://atos.cnj.jus.br/atos/detalhar/3429>>.

COSTA, Diego Carneiro. A responsabilidade civil pelos danos causados pela inteligência artificial nas hipóteses de discriminação algorítmica. **Revista Direito UNIFACS - Debate Virtual**, n. 272, 2023.

DANCEANU, Constanza Fernández. Systematic review of the causes of Latin American states' compliance with international human rights law. **Human Rights Quarterly**, v. 41, n. 3, p. 553-577, 2019.

DASTIN, Jeffrey. **Amazon Scraps Secret AI Recruiting Tool that Showed Bias against Women**, 2018. Disponível em: <<https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>>. Acesso em: 10 jun. 2023.

DINOS, Sokratis *et al.* **A systematic review of juries' assessment of rape victims: Do rape myths impact on juror decision-making?** **International Journal of Law, Crime and Justice**. [S.l: s.n.], 2015

DONATO, Helena; DONATO, Mariana. Stages for undertaking a systematic review. **Acta Medica Portuguesa**, v. 32, n. 3, p. 227-235, 2019.

DUARTE, Alan; NEGÓCIO, Ramon de Vasconcelos. Todos São Iguais Perante o Algoritmo? Uma Resposta Cultural do Direito à Discriminação Algorítmica. **Revista de Direito Público**, v. 18, n. 100, p. 218-244, 2021. Disponível em: <<http://orcid.org/0000-0003-4724-9658>>.

EUROPEAN UNION. **General Data Protection Regulation**. [S.l: s.n.], 2016.

FORTES, Pedro Rubim Borges. Paths to Digital Justice: Judicial Robots, Algorithmic Decision-Making, and Due Process. **Asian Journal of Law and Society**, v. 7, n. 3, p. 453-469, 2020.

FRANÇA NETTO, Milton Pereira de; EHRHARDT JÚNIOR, Marcos. Os riscos da discriminação algorítmica na utilização de aplicações de inteligência artificial no cenário brasileiro. **Revista Jurídica Luso Brasileira**, v. 8, n. 3, p. 1271-1318, 2022.

GABARDO, Emerson; FREITAS, Olga Lúcia C. de; VIANA, Ana Cristina Aguilar. The



digital divide in Brazil and the accessibility as a fundamental right. **Revista chilena de derecho y tecnología**, v. 11, 2022.

GOMES, Eduardo Biacchi; VAZ, Andréa Arruda; DIAS, Sandra Mara de Oliveira. Limites éticos para o uso da inteligência artificial no sistema de justiça brasileiro, de acordo com a Lei. 13.709 de 2018 (LGPD) e Resoluções 331 e 332 do Conselho Nacional de Justiça. **Revista Internacional Consinter de Direito**, v. 2, n. 13, 2021.

GUIMARÃES, Rodrigo Régner Chemim. A Inteligência Artificial e a disputa por diferentes caminhos em sua utilização preditiva no processo penal. **Revista Brasileira de Direito Processual Penal**, v. 5, n. 3, p. 1555–1588, 2019.

HACHEM, Daniel Wunder. São os direitos sociais “direitos públicos subjetivos”? Mitos e confusões na teoria dos direitos fundamentais. **Revista de Estudos Constitucionais, Hermenêutica e Teoria do Direito (RECHTD)**, São Leopoldo, v. 11, n. 3, p. 404-436, 2019.

HENRIQUES, Isabella Vieira Machado; SAMPAIO, Inês Vitorino. Discriminação Algorítmica e Inclusão em sistemas de Inteligência Artificial - Uma Reflexão sob a ótica dos Direitos da Criança no Ambiente Digital. **Direito Público**, v. 18, n. 100, p. 245–271, 2021.

HIROMI Saito, Vitória; SALGADO, Eneida Desiree. Privacidade e proteção de dados: por uma compreensão ampla do direito fundamental em face da sua multifuncionalidade. **International Journal of Digital Law**, v. 1, p. 117-137, 2020.

HORÁKA, Filip; LACKOB, David; KLOCEKB, Adam. Legal Consciousness: A Systematic Review of its Conceptualization and Measurement Methods. **Anuario de Psicología Jurídica**, n. 31, p. 9–34, 2021.

MARQUES, Fabíola; NETO MARTINEZ, Aldo Augusto. Vieses algorítmicos, direitos fundamentias e os sindicatos. **Revista Jurídica Luso Brasileira**, v. 8, n. 6, p. 707–729, 2022.

MEDEIROS, Alexandre Dimitri Moreira de; MOÇÃO, Josias. A questão ética da responsabilidade civil ao longo do ciclo de vida do sistema de inteligência artificial no processo seletivo de candidatos. 2021, Belo Horizonte: [s.n.], 2021.

MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. **Revista Direito Público**, v. 16, n. 90, p. 39–64, 2019.

Disponível em:

<https://www.academia.edu/42741206/Discriminação_Algorítmica_Conceito_Fundamento_Legal_e_Tipologia>.

MUGNAINI, Rogério; NOYONS, Ed; PACKER, Abel L. Fluxo de citações inter-nacional: fontes de informação para avaliação de impacto científico no Brasil Rogério. **6º Encontro Bras. Bibliometr. e Cientometria a ciência em rede**. 6º ed. Rio de Janeiro: UFRJ, 2018.

NATIONS, United. **International Covenant on Civil and Political Rights. General Assembly resolution 2200A**. [S.l: s.n.]. , 1966

_____. **Universal Declaration of Human Rights Preamble. United Nations General Assembly**. [S.l: s.n.]. , 1948



O'NEIL, Cathy. **Personality Tests Are Failing American Workers**, 2018. Disponível em: <<https://www.bloomberg.com/opinion/articles/2018-01-18/personality-tests-are-failing-american-workers#:~:text=Setting ADA violations aside for,with other types of assessments.>>. Acesso em: 10 jun. 2023.

OECD. **Recommendation of the Council on Artificial Intelligence**. . [S.l.: s.n.]. , 2019

OLIVEIRA, Thiago Dias. Content moderation technologies: Applying human rights standards to protect freedom of expression. **Human Rights Law Review**, v. 20, n. 4, p. 607–640, 2020.

PAEZ, Arsenio. Gray literature: An important resource in systematic reviews. **Journal of Evidence-Based Medicine**, v. 10, n. 3, p. 233–240, 2017.

PETERS, Micah D. J *et al.* **The Joanna Briggs Institute Reviewers ' Manual 2015 Methodology for JBI Scoping Reviews**. Adelaide: The Joanna Briggs Institute., 2015.

PINCUS, Fred L. From Individual to Structural Discrimination. In: PINCUS, Fred L; EHRLICH, Howard J (Org.). . **Race Ethn. Confl. Contend. Views Prejud. Discrim. Ethnoviolence**. New York: Routledge, 2018.

PURVES, Duncan; DAVIS, Jeremy. **Criminal justice algorithms: Being race-neutral doesn't mean race-blind**, 2022. Disponível em: <<https://theconversation.com/criminal-justice-algorithms-being-race-neutral-doesnt-mean-race-blind-177120>>.

PUSCHEL, André Felipe Silva; RODRIGUES, Roberto Tessis; VALLE, Vivian Cristina Lima López. The ethical dilemma of algorithmic decision in Public Administration. **A e C - Revista de Direito Administrativo e Constitucional**, v. 22, n. 90, p. 207–226, 2022.

REQUIÃO, Maurício; COSTA, Diego Carneiro. Discriminação algorítmica : ações afirmativas como estratégia de combate. **Civilistica**, v. 11, n. 3, p. 1–24, 2022.

REYNA, Justo; GABARDO, Emerson; SANTOS, Fabio de Sousa. Electronic government, digital invisibility and fundamental social rights. **Sequência (Florianópolis)**, n. 85, p. 30–50, 2020.

RIBEIRO, Ricardo Silveira. Inteligência artificial, direito e equidade algorítmica: discriminações sociais em modelos de machine learning para a tomada de decisão. **Revista de Informação Legislativa**, v. 59, n. 236, p. 29–53, 2022.

RIOS, Roger Raupp. Proteção de direitos LGBTQIA+ no Direito brasileiro: momentos e descompassos jurídicos e políticos. **Revista de Investigações Constitucionais**. v. 9, n. 3, p. 659-680, 2022.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. Harlow: Pearson, 2021.

SANTOS, Fábio de Sousa. Governo eletrônico, burocracia e controle: o uso da inteligência artificial preditiva na aplicação da Lei 13.655. de 2018. In: ZOCKUN, Maurício; GABARDO, Emerson (Coords.). **O Direito Administrativo do Pós-crise**. Curitiba: Íthala, 2021a.

_____. Racismo estrutural no Brasil e interpretação dos conceitos jurídicos indeterminados. In: ANDRADE, Giulia de Rossi; SAIKALI, Lucas Bossoni (Coords.). **Eficiência**,

subsidiariedade, interesse público e novas tecnologias. Curitiba: Íthala, 2021b.

SARLET, Ingo Wolfgang. **A Eficácia dos Direitos Fundamentais: uma Teoria Geral dos Direitos Fundamentais na Perspectiva Constitucional.** Porto Alegre: Livraria do Advogado, 2018.

_____. Fundamentos Constitucionais: o Direito fundamental à proteção de dados. In: DONEDA, Danilo et al. (Org.). **Tratado de proteção de dados pessoais.** Rio de Janeiro: Editora Forense, 2021.

SCHIER, Adriana da C. Ricardo; MAKSYM, Cristina Borges R.; MOTA, Vitória D. A urgência da regulação e do fomento da inteligência artificial à luz do princípio da precaução e do desenvolvimento sustentável. **International Journal of Digital Law.** v. 2, n. 3, p. 133-152, 2021.

SCHNELL, Joshua D. Web of Science: The First Citation Index for Data Analytics and Scientometrics. In: CANTÚ-ORTIZ, Francisco J. (Org.). **Res. Anal. Boost. Univ. Product. Compet. through Sci.** Boca Raton: Taylor & Francis, p. 15-29, 2018.

SCHOTTEN, Michiel *et al.* A Brief History of Scopus: The World's Largest Abstract and Citation Database of Scientific Literature. In: CANTÚ-ORTIZ, Francisco J. (Org.). **Res. Anal. Boost. Univ. Product. Compet. through Sci.** Boca Raton: Taylor & Francis, 2018. p. 289.

SOUSA, Thanderson Pereira; CRISTÓVAM, José Sérgio da Silva; MACHADO, Raquel Cavalcanti Ramos. Constitucionalismo e administração pública digitais: inovação tecnológica e políticas públicas para o desenvolvimento do Brasil. **Revista Brasileira de Políticas Públicas,** v. 12, n. 2, p. 178-196, 2022.

TOBBIN, Raissa Arantes; CARDIN, Valéria Silva Galdino. Biohacking and Ciborguism: Human Improvement in Light of Personality Rights. **Revista Opinião Jurídica,** v. 20, n. 35, p. 110-138, 2022.

TONIAZZO, Daniela Wendt; BARBOSA, Tales Schmidke; RUARO, Regina Linden. O Direito à explicação nas decisões automatizadas: uma abordagem comparativa entre o ordenamento brasileiro e europeu. **Revista Internacional Consinter de Direito,** v. 2, n. 13, 2021.

VEJA. **Decolar é multada em R\$ 7,5 milhões por diferenciar preços a consumidor,** 2018. Disponível em: <[VIGDOR, Neil. **Apple card investigated after gender discrimination complaints,** 2019. Disponível em: <<https://www.nytimes.com/2019/11/10/business/apple-credit-card-investigation.html>>](https://veja.abril.com.br/economia/decolar-e-multada-em-r-75-milhoes-por-diferenciar-precos-a-consumidor#:~:text=Em%20fevereiro%20o%20Minist%C3%A9rio%20P%C3%BAblico,que%20a%20compra%20era%20feita.>.>></p></div><div data-bbox=)

VINCENT, James. **Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day - The Verge,** 2016. Disponível em: <<https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>>

WERMUTH, Maiquel Ângelo Dezordi; CARDIN, Valéria Silva Galdino; WOLOWSKI, Matheus Ribeiro de Oliveira. Biopolítica E Novas Tecnologias: Direitos Humanos Sob Ameaça? **Rei - Revista Estudos Institucionais**, v. 7, n. 1, p. 276–296, 2021.

WIMMER, Miriam; DONEDA, Danilo. Carta da Editora e do Editor - Dossiê Temático “Inteligência Artificial, Ética e Epistemologia”. **Direito Público**, v. 18, n. 100, p. 7–17, 2021.

Sobre os(as) autores(as):

Nilton Sainz | *E-mail*: sainznilton@gmail.com

Doutorando em Ciência Política (UFPR), Mestre em Ciência Política (UFPEL), Pós-graduando em Data Science & Big Data (UFPR), Bacharel em Ciências Sociais (UFPEL), Graduando em Direito (PUCPR)

Emerson Gabardo | *E-mail*: emerson.gabardo@pucpr.br

Doutor em Direito do Estado (UFPR), Professor Titular de Direito Administrativo da PUCPR e Professor Associado de Direito Administrativo da UFPR. Pós-doutorado em Direito Público Comparado (Fordham University - EUA) e Ciência Política (University of California, UCI - EUA)

Natália Ongaratto | *E-mail*: natalia.ongaratto@gmail.com

Bacharela em Direito (FURG), Pós-graduada em Prática Jurídica Social (FURG) e Pós-graduanda na Fundação Escola do Ministério Público do Estado do Paraná (FEMPAR).

Data de submissão: 02 de julho de 2023.

Data da Triagem de Diretrizes: 15 de setembro de 2023.

Data da Triagem de Qualidade: 02 de dezembro de 2023.

Data de Envio para Avaliação: 26 de fevereiro de 2024.

Data da Primeira Avaliação: 14 de março de 2024.

Data da Segunda Avaliação: 10 de junho de 2024.

Data do aceite: 15 de junho de 2024

